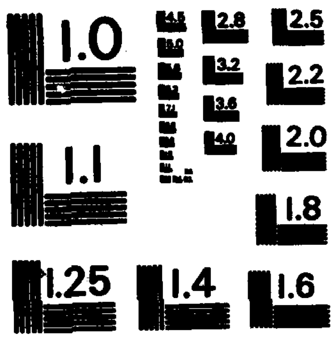
AD-A122 082 A STATISTICAL VALIDATION REPORT OF MARINE OPERATIONAL V/STOL ENVIRONMENT SIMULATION (MOVES)(U) NAVAL WEAPONS ENGINEERING SUPPORT ACTIVITY WASHINGTON DC W P HENKE 1/1

UNCLASSIFIED JUN 81 NAVWESA-TM-81-1 F/G 1/3 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A122082

(4)

A STATISTICAL VALIDATION REPORT

OF

MARINE OPERATIONAL V/STOL

ENVIRONMENT SIMULATION

(MOVES)

JUNE 1981

WILLIAM P. HENKE

OPERATIONS RESEARCH DIVISION
PRODUCTION DATA SYSTEMS DEPARTMENT
NAVAL WEAPONS ENGINEERING SUPPORT ACTIVITY
WASHINGTON NAVY YARD
WASHINGTON, D.C. 20374

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TM-81-1	2. GOVT ACCESSION NO. AD-A122 082	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Report on the Statistical Validation of Marine Operational V/STOL Environment Simulation (MOVES)		5. TYPE OF REPORT & PERIOD COVERED Final
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) William P. Henke		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Weapons Engineering Support Activity (ESA-33D) WNY, Washington, D.C. 20374		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Weapons Engineering Support Activity (ESA-33D) WNY, Washington, D.C. 20374		12. REPORT DATE June 1981
		13. NUMBER OF PAGES 24
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

MOVES, Simulation, Statistics, Validation, Computer models, computer simulation, Harrier, Modeling, Models

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report encompasses some methods applied when validating the MOVES model for experimentation under various U.S. Marine AV-8A Harrier squadron operating scenarios. The Statistical methods represented in the report are those that were applied for the U.S. Marine Users of MOVES. These statistical methods are the t-test, F-test, Confidence Intervals, Normal Distribution Plot, and the Wilcoxon-Mann-Whitney test. These methods will be used by the U.S. and U.K. when validating for the AV-8B and Sea Harrier MOVES.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AV	Attack V/STOL (Vertical Short Take-Off Landing)
BRN	British Royal Navy
MCAS	Marine Corps Air Station
NAVWESA	Naval Weapons Engineering Support Activity
RN	Royal Navy
S.K.	Sea-King (Helicopter)
UK	United Kingdom
UK MOVES	United Kingdom Maritime Operational V/STOL Environment Simulation
US	United States
US MOVES	United States Marine Operational V/STOL Environment Simulation

FOREWORD

This Technical Memorandum presents a report for reference during the process of validating the UK MOVES and US MOVES models. Validation is a process in which an inference is accepted that a simulator is a correct (valid) representation of a real-world situation. If a model has not been recently validated, it should not be used since the decision maker could not depend on its results. The driving force in the use of MOVES while experimenting with alternative solutions to problems, was that it be validated prior to running each particular set of experiments. The squadron/detachment was simulated and validated in their current state to guarantee to the users of MOVES that the model could accurately predict. Experiments were then run with a high degree of confidence put on the predictions.

The MOVES model is a computerized simulation model, and has been heavily used by the US Marine Corps, and UK Royal Navy in "try-before-buy" exercises, to simulate AV-8A Harrier and Sea Harrier operations, maintenance and supply support. By simulating via use of the validated MOVES model, problems and mistakes have been avoided, saving time, money and the inconvenience of living with expensive and undesirable situations that could have resulted from untried real-world actions.

This Statistical Validation report can be used in validating other simulation models, but should not be viewed as a comprehensive statistical validation text. This report illustrates those items that have been satisfactorily used and discussed with the U.S. and U.K. users of MOVES.

SUMMARY

The problem under study was how to convince the Harrier aircraft decision makers, e.g., squadron operator, NAVAIR and DOD management, that the MOVES model was worth the trouble and cost of using it. Many times the question posed to NAVWESA was, "After we describe operating logic and data to you, how do we know all our effort was worth it, i.e., how do we know the model is any good?" This concern was genuine, since the elements and data addressed by MOVES are many in number and broad in scope, and many people have spent a lot of time and effort into the construction of simulation models that proved worthless. The approach to answer the question was by proving to these decision makers that MOVES could accurately predict the consequences of experimenting on simulated squadron activities, thus avoiding the time and trouble of experimenting on the actual real-world process itself. This proof was established by use of the methods presented in this validation plan during the initial validation of MOVES in 1975. That is, in 1975, flight operations, maintenance and supply support of Squadron VMA-542 were simulated using the MOVES model. The output of MOVES for 3 consecutive months agreed (was valid) with the real-world output of the Squadron. Since then, the model has been continuously validated against each squdaron prior to its use for experimentation on that squadron, again using the techniques presented in this validation plan..

The MOVES model, a computerized simulation of Harrier squadron flight operations, maintenance and supply support, was conceived, designed and developed by NAVWESA for PMA-257 (NAVAIR AV Project Manager). Its use has been input to changes in policy, including stretching of intervals between phased maintenance, use of the team maintenance concept, and cannibalization. In 1978, the British Government, under contract with the U.S. Government, adopted MOVES for use as their Royal Navy Sea Harrier and Sea King Helicopter simulation model, via support by NAVWESA. MOVES is a highly effective tool for use in preventing problems vice curing them, in the "try-before-buy" mode exercised by the U.S. and U.K.

As a result of the use of this report to validate the MOVES model, users in the U.S. and U.K. exhibit a high acceptance of MOVES. This report should be used as reference when validating MOVES if MOVES were tailored for non-AV aircraft, or for validating other simulation models.

PURPOSE

The purpose of this report is to provide a reference for use when validating the UK MOVES model and future US MOVES models.

PROBLEM

In simulation work, the most important and difficult problem to solve is that the representation (model) of the real-world must be legitimate (valid) and to the satisfaction of the model user(s). It then behooves the simulation expert to establish for the user, a high level of confidence in the predictive powers of the model and its valid use for experimentation. This problem was kept on the surface during the design, development and implementation of the US MOVES model. To solve the problem, high levels of confidence were established in the US MOVES AV-8A model by continual use of the methods described in this Statistical Validation Plan. That is, prior to each AV-8A experiment, MOVES was validated for that particular squadron/detachment, operating in their current mode. Then experiments were run, and exhaustive analyses performed on the interaction of critical elements within the model, as well as on the predictions output by MOVES.

VALIDATION

Validation is the process of building an acceptable level of confidence that an inference about a simulated process is a correct (valid) inference for the real-world process. The validation of the MOVES model is considered complete when an agreement (objective or subjective) is reached between the behavior of the simulation (MOVES model) and a real-world system; the model is then valid for use. The simulator (MOVES) is defined as a symbolic or numerical abstraction of the real-world process under study, and is not the real-world process itself. The goal of simulation using the MOVES simulator is to learn something about the anticipated real-world process vice experimenting on the real-world process itself.

Since simulations tend to become far more complex than other Management Science models, validation is necessary to instill an acceptable level of confidence when simulating a real-world process. Simulators allow the decision maker to include many different parts and processes in one model, and allow them to interact in non-linear, non-steady-state modes. The worth of a simulation model with untested, untestable, or refuted assumptions is questionable.

Seldom will validation result in an absolute proof that the simulator is a "true" model of the real process. The simulator produces some output from each run; this output then needs validation. Validation should be done under various conditions of land and sea basings. Many approaches to the proof of validation exist, each of which may increase or decrease the confidence in the simulation output. These approaches include:

a. Various statistical tests on simulator output versus real-world performance, including confidence intervals, tests of means and variances, analysis of variance, regression, factor analysis, spectral analysis and auto-correlation, chi-square, and non-parametric tests. See Appendices A,B,C,D for statistical discussions.

b. Special data collection efforts in which data not normally collected are obtained from the real-world for comparison with data output at various steps within the simulator. This important area of validation should be well analyzed prior to application, since it involves additional data collection.

c. Field tests in which a process is placed in an operational situation and performance is measured prior to actual implementation. This process is then monitored in actual operation, comparing results with the MOVES internal output such as queue congestion.

d. Complementary research to determine "why" situations occur, and possible ways to improve the situation.

The US MOVES AV-8A model was validated initially in 1975, and was continually validated before each series of simulation experiments. The US MOVES AV-8B and UK MOVES models, although not validated as of this date, have been run using best estimates of input data and logic describing the expected real-world. As data and information concerning squadron operations evolve, both MOVES models should be run in conjunction with operating the Harriers on land or sea. These two (MOVES and squadron operations) will complement each other, resulting in shorter solutions to operating problems. Pre-validation runs of the MOVES models should be made in various configurations of land or sea basings, along with analyses as described in a,b,c, and d above. This will increase confidence in the models' predictive powers during and after validation, enabling decision makers to gain insight into solutions of potential problems.

Validation implies use of the model. The "users" of the model are land and sea squadron managers and operators, along with others concerned with policies or other decisions regarding the operation of the Harrier. If these "users" are not motivated to use the model, the model is less than useful; it becomes an academic exercise.

BACKGROUND

The validation process of the US MOVES model initially took place in August 1975, using VMA-542 AV-8A squadron operating at MCAS Cherry Point, North Carolina. Because MOVES output had been continually validated, those charged with making decisions about the AV-8A used it to develop policy in the peacetime and wartime areas of phase-maintenance, cannibalization, flying programs, on-deck maintenance, and maintenance work-center configuration. The users of the AV-8A US MOVES model bear witness to a high degree of confidence in the predictive powers of the model. This validated model is presently under transition from the AV-8A configuration to the advanced AV-8B configuration.

As a result of the confidence generated by the US MOVES validation and use, in 1978 the BRN contacted the US Navy to have NAVWESA modify the US MOVES model to represent the British environment in which the Royal Navy Sea Harriers would operate aboard ship in conjunction with the Sea King helicopters. This work resulted in the creation of the UK MOVES model. Many experiments have been run on this UK model, assisting the BRN in the composition of wartime/peacetime flying scenarios, along with policies involving maintenance, supply support, and the interaction between the Sea Harrier and Sea King.

The MOVES Model

MOVES is a computerized simulation of the operations, maintenance, supply and logistics support of Harrier aircraft squadrons/detachments. The MOVES model predicts the affect that various combinations of flying programs, maintenance programs, supply and logistic support have on the capability of the squadron/detachment to meet its mission. Some elements that the model predicts as a result of each simulation include; operational readiness, flying hour program achievement, sorties flown and lost, non-operational-ready due to supply and maintenance, downtime due to scheduled and unscheduled maintenance, direct-maintenance-manhours summary, cannibalization summary, supply profile showing each part ordered from supply, number of times the part was in supply and number of times the part had to be cannibalized. In addition, MOVES lists a complete profile of maintenance work center congestion along with utilization of support equipment. Two versions of MOVES exist;

1) the US MOVES supporting the US Marine Corps in their planning for operating the AV-8A Harriers, and design and operation of the AV-8B Harriers; 2) the UK MOVES supporting the British Royal Navy (BRN) Sea Harrier operations. Both models were conceived, designed, and developed by the Program Analysis Department (ESA-3) of the Naval Weapons Engineering Support Activity (NWESA).

The shelves are full of computerized, often costly, simulation models not in use due to reasons including: the inability to validate; the inability to obtain useful input data; poor and untimely communication between model developers and model users; inappropriate level of detail; too much generalization; model designers unfamiliar with the real-world being modeled; inappropriate output statistics for use in decision making; and the length of time for design/development/validation, precluding timely input of model results in decision making. The MOVES model is not one of these models.

CONCLUSIONS AND RECOMMENDATIONS

The use of the methods presented in this report is valuable to the process of validation of the MOVES model, a large and complex simulation model. Through this comprehensive validation via use of these methods, decision makers feel a high degree of confidence with the predictive results of the MOVES model. The results of applying these scientific methods are input to the human decision process in proving comprehensive validation, leading to extensive use of the simulation model.

The methods presented in this report should be used as a reference when validating other simulation models; large or small, complex or simple.

Appendix A
Sampling Distributions

A distribution which shows the probabilities of obtaining different values of $\overline{x}$ (sample mean) is called a theoretical sampling distribution of $\overline{x}$.

Two important theorems are related to theoretical sampling distributions:

1) If random samples of size n are taken from a population which has the mean μ and standard deviation σ, the theoretical sampling distribution of $\overline{x}$ has the mean μ and the standard deviation $\sigma/(n)^{\frac{1}{2}}$.

This theorem states that the mean of the sampling distribution of $\overline{x}$ equals the population mean. Also, $\sigma/(n)^{\frac{1}{2}}$ (called the standard error of the mean) is the standard deviation of the theoretical sampling distribution.

The standard error of the mean plays an important role in inductive statistics, because it measures the variation of the theoretical sampling distribution of $\overline{x}$. In other words, it tells how much the sample means can be expected to vary from sample to sample. The standard error of the mean decreases as n increases. That is, when n is large, a sample mean will yield a more reliable estimate of μ than when n is small.

2) If n is large (≥ 30) and is composed of the sum of independent identically distributed random variables, the theoretical sampling distribution of $\overline{x}$ can be approximated very closely with a <u>normal curve</u>, regardless of the shape of the population distribution. This is called the Central Limit theorem, and permits the calculations of the probabilities of obtaining various values of $\overline{x}$. Note that this theorem contains no specification about the shape of the distribution of the population.

Utilizing these 2 theorems, it is possible to construct a statistical confidence interval about the mean, using the sampling mean $\overline{x}$, and σ. Although $\overline{x}$ estimates the mean μ, the value of the population standard deviation, σ, is not known. Therefore, σ must be replaced with an estimate, i.e., with the sample standard deviation s:

$$s = \left[\frac{n\Sigma x^2 - (\Sigma x)^2}{n\ (n-1)} \right], \text{ where } \Sigma = \text{summation.}$$

Sample standard deviations can be expected to provide close estimates of σ when $n \geq 30$. In computing confidence intervals around the mean, for this case ($n \geq 30$) it is necessary to refer to a table of Normal Curve Areas; for $n < 30$, it is necessary to refer to a table of "t" values. In dealing with "t" values, it is assumed that the sample is random and the population can be approximated closely with a normal curve.

Appendix B

Student t-test, F-test, and Confidence Intervals

Student t-test

An important part of the validation process is to compare the outputs of the real system with those generated by its model representation. A method of performing this comparison is by using the Student t-test, a statistical tool generally applied when the number of data points available is less than 30. The Student t-test distribution on which the t-test is based, is peaked in the center and has higher tails than does the normal distribution. Specifically, two procedures using the t distribution can be utilized to 1) compare the means of 2 samples, and 2) construct confidence intervals about those mean values. These two procedures are now discussed in greater detail.

Procedure 1:

Assume that 2 populations with true but unknown means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 respectively are the subject of analysis. Then use of the t-test requires the following 2 assumptions: 1) the observations made from each population are random, independent, and drawn from an approximately normal population; 2) the variances of the 2 populations are equal, that is $\sigma_1^2 = \sigma_2^2$.

Each set of observations drawn from the population is called a sample. If 2 samples, 1 from each population are taken, and the mean values of the samples are m_1 , m_2 respectively, then the <u>first procedure</u> seeks to answer the following question: Is the difference $m_1 - m_2$ attributable to a population difference $\mu_1-\mu_2$, or may it be random variation from a single population mean?

Procedure 2:

Since the sample mean m_1 (m_2) is only an approximation of the true mean μ_1 (μ_2) of the first (second) population, the <u>second procedure</u> is designed to answer this question: In the long run, between what 2 values will the true (but unknown) mean lie a certain percent of the time? The phrase "a certain percent of the time" is defined in terms of a number between 0 and 1.0, and is called the confidence level. A confidence level is selected, the sample is drawn, then the lower and upper values (called confidence limits) which could include the unknown mean will be calculated. This interval between the upper and lower confidence limits is known as a confidence interval. If the sampling is repeated indefinitely, with each sample leading to a new confidence interval (a new interval estimate), then in 95% of the samples, the interval will cover

the true population mean. If one makes a practice of sampling, and if for each sample one states that the true mean lies within the confidence interval, 95% of these statements will be correct; 5% will be wrong.

The comparison of means technique is illustrated as follows. Let an AV-8A/B Harrier squadron operating in a ship or land based environment be the real system, and let the USMOVES simulation model be the computer representation of that squadron. Assume that squadron operations have reached a near-stable state and that the quantity of sorties flown per week is the subject of validation. In terms of the above description, squadron summary reports represent observations of the real system, and the replication of simulation experiments represents observations of the simulation model. During the time period being studied, the question is the following: Does the simulation model adequately represent squadron operations in terms of sorties flown?

Suppose that a one week period of squadron operations is simulated and that this experiment is replicated with different random seeds 4 more times giving a sample of size 5. Replications in this case are random independent observations from the population of all replications for this experiment. This sample of size 5 is to be compared with a sample of 8 squadron summary reports, each summarizing a period of 1 week. Note that consecutive 1 week summaries are not necessarily independent since there are cases where the operations of 1 week heavily influence those of the following week. Although a sample of completely independent summaries is nearly impossible to select, this problem should still be considered. Unless evidence to the contrary is available, both populations are assumed to be normal. The assumption of equal variances, if in doubt however, can be statistically checked as will be demonstrated further. Suppose now that the following 2 samplings (1 from the squadron and 1 from the model) representing sorties flown per week were obtained:

	Sampling 1 Squadron Summary		Sampling 2 Model Output
	48		56
	53		63
	51		54
	60		58
	55		56
	57	TOTAL	287
	53		
	53		
TOTAL	430		

Now calculate the mean of each sample by computing mean
$= \frac{\text{sum of sorties in sample i}}{\text{size of sample i}}$

mean sample 1 = 430/8 = 53.75 = $\bar{X}_1$
mean sample 2 = 287/5 = 57.4 = $\bar{X}_2$

The Student t-test can now be applied to examine the difference $\bar{X}_1 - \bar{X}_2$. If the difference is small enough, this will indicate that both means are considered to be estimates of the same population mean, and therefore the simulation model seems to adequately represent the flying program of the squadron. If the difference is not small enough, the indication is that the model and squadron means represent different populations; if so, modification of model logic, or additional data analysis (or both) may be necessary. The t-test is performed by calculating the statistic:

$$t = (\bar{X}_1 - \bar{X}_2)\left[\frac{n_1 n_2(n_1+n_2-2)}{(n_1+n_2)\,\Sigma X^2}\right]^{\frac{1}{2}}$$ where $\bar{X}_1$, $\bar{X}_2$ are

sample means, n_1 and n_2 the sample sizes of samples 1 and 2 respectively, and ΣX^2 is the pooled sum of squares for the 2 samples, which is caluclated by subtracting the sample mean from the observed sample value, squaring the result and summing over all observations.

$$\Sigma X^2 = (48-53.75)^2 + (53-53.75)^2 + \cdots + (53-53.75)^2 + (56-57.4)^2 = (63-57.4)^2 + \cdots (56-57.4)^2$$

$$= 93.5 + 47.18 = 140.7$$

setting $n_1=8, n_2=5$, $t = (53.75 - 57.4)\left[\frac{8(5)\ (8+5-2)}{(8+5)\ (140.7)}\right]^{\frac{1}{2}}$

$$= 3.65\left[\frac{40\ (11)}{13\ (140.7)}\right]^{\frac{1}{2}} = 1.79$$

This t-value is now compared to a value obtained from page B-6 of values for the t-distribution for various confidence levels. Assume a confidence level of 95%. The table value obtained for 13-2 = 11 degrees of freedom is 2.201 (2 is subtracted from the total of 13 since 2 mean values were calculated). Since the calculated t-value 1.79 is less than the table value of 2.201, the observed sortie average difference is not significant; i.e., the model sortie output average adequately represents the squadron sortie output average. If the calculated t-value exceeded 2.201 then model logic/data could have required more analysis.

F-test

If the assumption of equal population variances is in doubt, another statistical test called the F-test may be used. The confidence level concept forms the basis of the F-test just as it did in the t-test. In general the F-test compares the variances of samples drawn from 2 populations and if the ratio of the larger variance divided by the smaller exceeds a table value chosen at a specified confidence level, then the hypothesis that the population variances are equal would be rejected. To apply the test to the 2 samples listed above, calculate the sample variances as follows:

$$\text{variance}_i = \frac{\text{pooled sum squares sample i}}{\text{(size sample i) - 1.}}$$

$$\text{variance}_1 = ((48 - 53.75)^2 + (53 - 53.75)^2 + .. + (53-53.75)^2)/(8-1)$$

$$= 93.5 / 7$$

$$= 13.36$$

$$\text{variance}_2 = ((56-57.4)^2 + (63-57.4)^2 + .. (56-57.4)^2)/(5-1)$$

$$= 47.18 / 4$$

$$= 11.80$$

The computed F-value is $F = \dfrac{\text{variance}_1}{\text{variance}_2}$

$$= 13.36 / 11.80$$

$$= 1.13$$

The table value as shown on page B-7 for F assuming a confidence levelof 95% for 8-1 = 7 and 5-1 = 4 degrees of freedom respectively is 6.09. Since the calculated F-value is less than the table value, based on the sample data listed above, the hypothesis that populations have equal variances cannot be rejected. If the calculated value had exceeded the table value, then application of the t-test previously would have been questionable, and other statistical tools/analyses would be required.

Confidence Intervals

The most important aspect involved in the application of confidence intervals is not in the statistics of the exercise, but in the agreement between the population and the sample. Claims that confidence interval estimates apply to the population that was actually sampled, rest on the judgement of the investigator. Careful investigators take pains to describe any relevant characteristics of their data in order that others can envisage the nature of the sampled population.

Consider a confidence interval, at a probability (confidence) level of 95%. Before the random sample is drawn, the confidence statement is, "the probability is 95% that the interval to be constructed would include the true (but unknown) population mean". After the sample is drawn, the confidence statment is true or false; it is incorrect to say "the probability is 95% that the population mean lies between the lower and upper value of the sample just taken". In a specific application, we do not know if our confidence statement is actually one of the 95% correct, or one of the 5% that are wrong.

If the sampling is repeated indefinitely with each sample leading to a new confidence interval (a new interval estimate), then in 95% of the samples, the interval will cover the true population average. If one makes a practice of sampling, and if for each sample, states that the true mean lies within the confidence interval, 95% of the statements will be correct.

The uncertainty comes from the sampling process. Each sample specifies an interval estimate. Whether or not each interval happens to include the true population mean is a risk (probability). This probability is not of the true mean lying in the interval, because the true mean is unknown and fixed; thus cannot have a distribution. The risk is the probability of the interval, (the random variable) containing the true value. The confidence statement concerns the population mean; it does not concern the mean in other samples to be drawn.

TABLE B. TABLE OF CRITICAL VALUES OF t*

df	Level of significance for one-tailed test					
	.10	.05	.025	.01	.005	.0005
	Level of significance for two-tailed test					
	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

* Table B is abridged from Table III of Fisher and Yates: *Statistical tables for biological, agricultural, and medical research*, published by Oliver and Boyd Ltd., Edinburgh, by permission of the authors and publishers.

PERCENTILES OF THE F DISTRIBUTION

$F_{.95}(n_1, n_2)$

n_1 = degrees of freedom for numerator

n_2 = degrees of freedom for denominator

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Adapted with permission from *Biometrika Tables for Statisticians*, Vol. I (2d ed.), edited by E. S. Pearson and H. O. Hartley, Copyright 1958, Cambridge University Press.

Appendix C
Wilcoxon-Mann-Whitney Test

Wilcoxon-Mann-Whitney Test

A most useful alternative to the t-test is the non-parametric Wilcoxon-Mann-Whitney test. This can be used to test if the average of two sets of independent samples differ. The procedure used is to:

a. Choose a significance level of the test. This is called α, and is equal to 1.00 - Confidence level.

b. Combine the observations from the two samples and rank them in order of increasing size from smallest to largest. Assign 1 to smallest, 2 to next smallest etc. In case of ties, average the rankings.

c. Let n_1 = smaller sample size, n_2 = larger sample size and N = total sample size.

d. Compute R_1 = sum of ranks for the smaller sample. If samples sizes are equal, use sum of ranks for either sample.

e. Compute $R_2 = n_1 (N + 1) - R_1$.

f. Find value in table on page C-3 or C-4 (Critical R Values) of R for α and (n_1,n_2). If either R_1 or R_2 is smaller than, or equal to, the table R for (n_1,n_2), the averages of the two samples differ. Otherwise, there is no reason to believe that the averages differ.

For example:

a. $\alpha = 1.00-.90=.10$, for 2-sided test

b.

Group A	Group B
50.5 (9)	57.0 (17)
37.5 (1)	52.0 (11)
49.8 (7)	51.0 (10)
56.0 (15.5)	44.2 (3)
42.0 (2)	55.0 (14)
56.0 (15.5)	62.0 (19)
50.0 (8)	59.0 (18)
54.0 (13)	45.2 (5)
48.0 (6)	53.5 (12)
	44.4 (4)

Numbers in parentheses indicate rankings.

c. $n_1 = 9$, $n_2 = 10$, $N = 19$

d. $R_1 = 77$

e. $R_2 = 9\ (20) - 77 = 103$

f. R for .10 and (9, 10) = 69. Since neither R_1 or R_2 are smaller than 69, there is no reason to believe that the averages of the two groups differ.

R(RANK) VALUES

CRITICAL VALUES OF SMALLER RANK SUM FOR THE WILCOXON-MANN-WHITNEY TEST

n_2	α for 2-Sided Test	α for 1-Sided Test	n_1 (Smaller Sample) 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	.20	.10		3	7																	
	.10	.05			6																	
	.05	.025																				
	.01	.005																				
4	.20	.10		3	7	13																
	.10	.05			6	11																
	.05	.025				10																
	.01	.005																				
5	.20	.10		4	8	14	20															
	.10	.05		3	7	12	19															
	.05	.025			6	11	17															
	.01	.005					15															
6	.20	.10		4	9	15	22	30														
	.10	.05		3	8	13	20	28														
	.05	.025			7	12	18	26														
	.01	.005				10	16	23														
7	.20	.10		4	10	16	23	32	41													
	.10	.05		3	8	14	21	29	39													
	.05	.025			7	13	20	27	36													
	.01	.005				10	16	24	32													
8	.20	.10		5	11	17	25	34	44	55												
	.10	.05		4	9	15	23	31	41	51												
	.05	.025		3	8	14	21	29	38	49												
	.01	.005				11	17	25	34	43												
9	.20	.10	1	5	11	19	27	36	46	58	70											
	.10	.05		4	9	16	24	33	43	54	66											
	.05	.025		3	8	14	22	31	40	51	62											
	.01	.005			6	11	18	26	35	45	56											
10	.20	.10	1	6	12	20	28	38	49	60	73	87										
	.10	.05		4	10	17	26	35	45	56	69	82										
	.05	.025		3	9	15	23	32	42	53	65	78										
	.01	.005			6	12	19	27	37	47	58	71										
11	.20	.10	1	6	13	21	30	40	51	63	76	91	106									
	.10	.05		4	11	18	27	37	47	59	72	86	100									
	.05	.025		3	9	16	24	34	44	55	68	81	96									
	.01	.005			6	12	20	28	38	49	61	73	87									
12	.20	.10	1	7	14	22	32	42	54	66	80	94	110	127								
	.10	.05		5	11	19	23	38	49	62	75	89	104	120								
	.05	.025		4	10	17	26	35	46	58	71	84	99	115								
	.01	.005			7	13	21	30	40	51	63	76	90	105								

13	.20	.10	1	7	15	23	33	44	56	69	83	98	114	131	149							
	.10	.05		5	12	20	30	40	52	64	78	92	108	125	142							
	.05	.025		4	10	18	27	37	48	60	73	88	103	119	136							
	.01	.005			7	14	22	31	41	53	65	79	93	109	125							
14	.20	.10	1	7	16	25	35	46	59	72	86	102	118	136	154	174						
	.10	.05		5	13	21	31	42	54	67	81	96	112	129	147	166						
	.05	.025		4	11	19	28	38	50	62	76	91	106	123	141	160						
	.01	.005			7	14	22	32	43	54	67	81	96	112	129	147						
15	.20	.10	1	8	16	26	37	48	61	75	90	106	123	141	159	179	200					
	.10	.05		6	13	22	33	44	56	69	84	99	116	133	152	171	192					
	.05	.025		4	11	20	29	40	52	65	79	94	110	127	145	164	184					
	.01	.005			8	15	23	33	44	56	69	84	99	115	133	151	171					
16	.20	.10	1	8	17	27	38	50	64	78	93	109	127	145	165	185	206	229				
	.10	.05		6	14	24	34	46	58	72	87	103	120	138	156	176	197	219				
	.05	.025		4	12	21	30	42	54	67	82	97	113	131	150	169	190	211				
	.01	.005			8	15	24	34	46	58	72	86	102	119	136	155	175	196				
17	.20	.10	1	9	18	28	40	52	66	31	97	113	131	150	170	190	212	235	259			
	.10	.05		6	15	25	35	47	61	75	90	106	123	142	161	182	203	225	249			
	.05	.025		5	12	21	32	43	56	70	84	100	117	135	154	174	195	217	240			
	.01	.005			8	16	25	36	47	60	74	89	105	122	140	159	180	201	223			
18	.20	.10	1	9	19	30	42	55	69	84	100	117	135	155	175	196	218	242	266	291		
	.10	.05		7	15	26	37	49	63	77	93	110	127	146	166	187	208	231	255	280		
	.05	.025		5	13	22	33	45	58	72	87	103	121	139	158	179	200	222	246	270		
	.01	.005			8	16	26	37	49	62	76	92	108	125	144	163	184	206	228	252		
19	.20	.10	2	10	20	31	43	57	71	87	103	121	139	159	180	202	224	248	273	299	325	
	.10	.05	1	7	16	27	38	51	65	30	96	113	131	150	171	192	214	237	262	287	313	
	.05	.025		5	13	23	34	46	60	74	90	107	124	143	163	182	205	228	252	277	303	
	.01	.005		3	9	17	27	38	50	34	78	94	111	129	147	168	189	210	234	258	283	
20	.20	.10	2	10	21	32	45	59	74	90	107	125	144	164	185	207	230	255	280	306	333	361
	.10	.05	1	7	17	28	40	53	67	83	99	117	135	155	175	197	220	243	268	294	320	348
	.05	.025		5	14	24	35	48	62	77	93	110	128	147	167	188	210	234	258	283	309	337
	.01	.005		3	9	18	28	89	52	66	81	97	114	132	151	172	193	215	239	263	289	315

For larger values of n_1 and n_2, critical values are given to a good approximation by the formula:

$$\frac{n_1}{2}(n_1 + n_2 + 1) - z\left\{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}\right\}^{1/2}$$

where $z = 1.28$ for $\alpha = .20$ (two-sided test)
$z = 1.64$ for $\alpha = .10$ "
$z = 1.96$ for $\alpha = .05$ "
$z = 2.58$ for $\alpha = .01$ "

Appendix D

Determining if observed data fit a Normal Curve.

There are a number of methods by which one can relatively quickly determine if observed data are normally distributed. This Appendix discusses three methods.

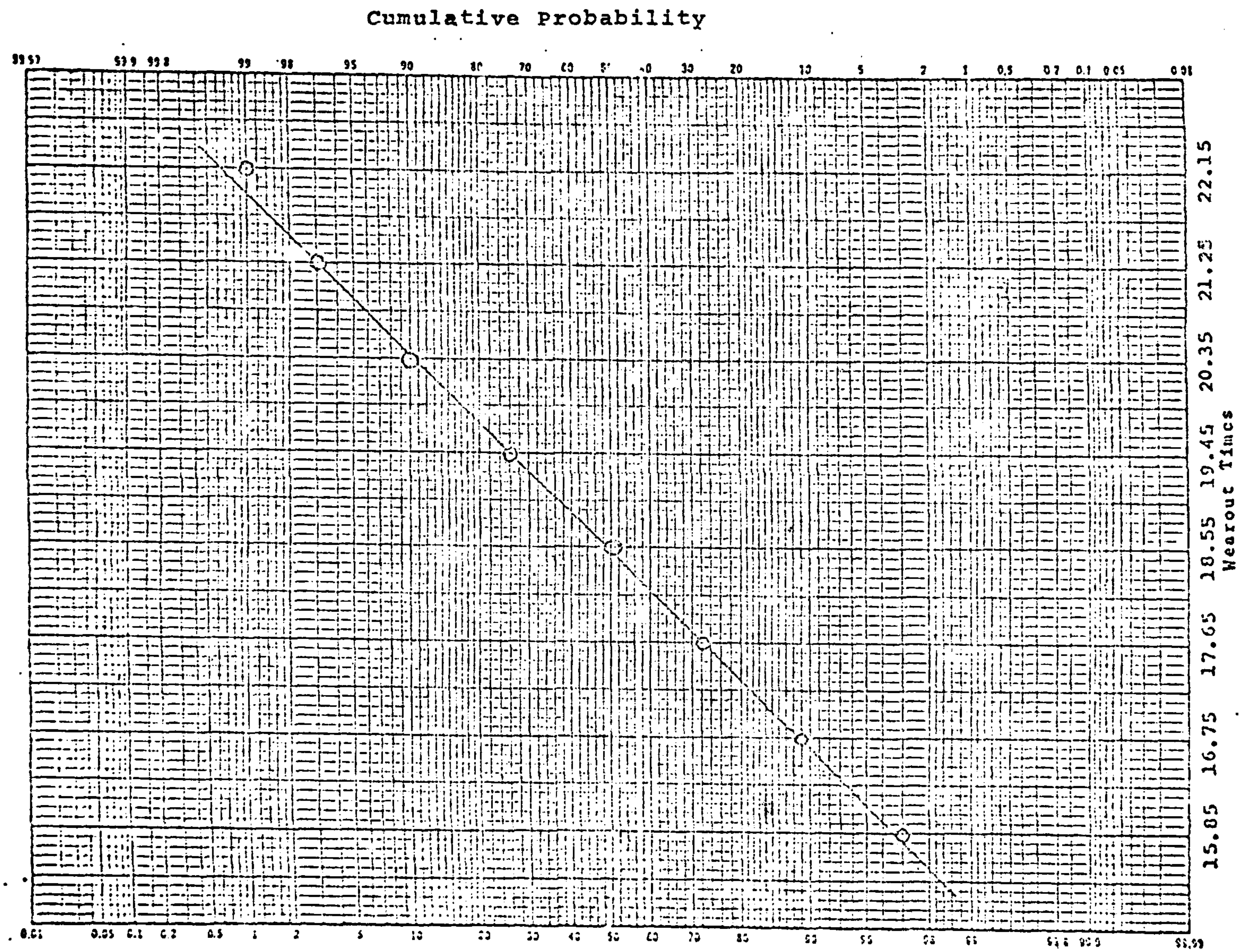
1) If the cumulative "less than" percentage distribution of a set of data which fits closely to a Normal Curve is plotted on probability paper, the points will lie on a straight line (or reasonably close to a straight line). For example, consider the following data distribution of wearout times of a particular mechanical assembly:

Class Limits; Wearout Times	Frequency
15.0 - 15.8	9
15.9 - 16.7	24
16.8 - 17.6	51
17.7 - 18.5	66
18.6 - 19.4	72
19.5 - 20.3	48
20.4 - 21.2	21
21.3 - 22.1	6
22.2 - 23.0	3
	300

Converting this information into a cumulative percentage distribution yields:

Class boundaries; Wearout Times	Cumulative Percentage	
less than 14.95	0	
less than 15.85	3	plot these points on probability paper as shown on page D-2
less than 16.75	11	
less than 17.65	28	
less than 18.55	50	
less than 19.45	74	
less than 20.35	90	
less than 21.25	97	
less than 22.15	99	
less than 23.05	100	

It seems reasonable to state that on the basis of the plot shown on page D-2, the points lie very close to a straight line; hence the original distribution of data can be approximated very closely with a Normal Curve. Note that


Cumulative Probability
Wearout Times
15.85
16.75
17.65
18.55
19.45
20.35
21.25
22.15

the first and last "less than" boundaries were not plotted, since 0 or 100 percent of the Normal Curve can never be reached.

A disadvantage of this method is that one must decide subjectively whether the points fall "reasonably close" to a straight line. This subjectivity can be eliminated by using a more precise method called the "Goodness of Fit" test to the data. This test is the Chi-Square (χ^2) test and statistically compares the frequencies from the original distribution to the expected frequencies from a theoretical normal curve. The explanation of the χ^2 test is not presented herein. The χ^2application in this case can be calculated by the interested reader to determine if the observed distribution constitutes a sample from a population having a Normal Distribution.

2) Another method of identifying a possible Normal Distribution, is if the mean=median=mode (or if they are reasonably close), then the observed distribution closely follows a Normal Distribution pattern. The mean is defined as the arithmetic mean. The median is the value that corresponds to the point which divides the distribution into two equal parts. The mode is the value which occurs most often.

3) A third method is to compute the mean and standard deviation of a set of representative data. If much more or much less (admittedly subjective) than 5% of the distribution lies outside the limits of mean $\pm$ 2 times standard deviation, the data probably follow a distribution other than the Normal.

Appendix E

Assumptions

An Assumption as related to the operation of the U.K. MOVES model is defined as one of the multifarious elements that is not explicitly addressed throughout each simulation. All other elements are explicitly addressed in the model and are defined as rules. This differentiation can be considered arbitrary; it is a convenient way of grouping these items, and is submitted with this plan for information purposes.

The list of assumptions include:

1. The methodology contained within the model of the real-world (including input data and output calculations) adequately describes the real-world situation, and this methodology remains intact when translated into the computer program (computerized).

2. The real-world situation being simulated is in a steady-state condition.

3. The form of the input data has been mathematically described, and is consistent with its use in the model.

4. Data at the component level are consistent with the same data at the parent system level.

5. Tractors, tow-arms, mechanical handlers and lifts never fail to operate (100% reliable).

6. The Sea King is 100% reliable.

7. Combat and non-combat damage, personnel casualties (combat and non-combat), weather, fueling and re-fueling, fires, do not affect the operation of the aircraft or squadron.

8. Skill level within the supervisory and non-supervisory ranks remain constant throughout each simulation.

9. Time spent on breaks from work (tea, lunch, etc.) when on-shift, is included in repair time.

10. Men are in proper location when needed. The lift is always in proper position when needed.

11. When a part is required and is in stock, the part is obtained instantaneously.

12. Time-to-next-failure and supply-delay-times follow an exponential distribution; maintenance-times are the average repair times. (In the U.S. MOVES model, log-normal repair times are used.)